2009

Determining The Effect Of Speaker's Gender And Speech Synthesis On Callsign Acquisition Test (CAT) Results

Jasminkumar B. Soni
Wright State University

Determining The Effect Of Speaker's Gender And Speech Synthesis On Callsign Acquisition Test (CAT) Results.

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Science in Engineering

By

Jasminkumar Soni
B.E., Saurashtra University, India, 2005

2009
Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

March 05, 2009

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY JASMINKUMAR BIPINCHANDRA SONI ENTITLED Determining The Effect Of Speaker's Gender And Speech Synthesis On Callsign Acquisition Test (CAT) Results BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Engineering.

Misty Blue, Ph.D.
Thesis Director

S.Narayanan, Ph.D.
Department Chair

Committee on
Final Examination

Misty Blue, Ph.D.

Yan Liu, Ph.D.

Blair A. Rowley, Ph.D.

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies

ABSTRACT

Soni Jasminkumar B. M.S.Egr., Department of Biomedical, Industrial and Human Factor Engineering, Wright State University, 2009. Determining The Effect Of Speaker's Gender And Speech Synthesis On Callsign Acquisition Test (CAT) Results.

Effective and efficient speech communication is one of the leading factors for success of battlefield operation. With the increases in the levels of gender diversity in military services, it is important to assess the effectiveness of voice for both genders in communication systems. The purpose of this research study was to determine the effect of the speaker's voice (male and female) on the speech intelligibility (SI) performance of the Callsign Acquisition Test (CAT). In addition, the effects of synthesized speech were evaluated. The CAT test is a new SI test that has been developed for military use. A group of 21 listeners with normal hearing participated in the study. Each participant listened to four different lists of CAT (male and female natural recorded speech, and male and female synthetic speech) at two signal-to-noise ratios. White noise was used as a masking noise and various speech files were mixed at signal-to-noise ratios -12 dB and -15 dB. Each wordlist was played at 50dB and 53dB mixed with white noise at 65dB. Each listener participated in a total of 8 tests presented in a random fashion. Testing was performed in a sound treated booth with loud speakers. Test results demonstrated that male speech and natural voice have higher SI results than female speech and synthetic voice respectively. Also statistical analysis concluded that female speech, -15 dB SNR, synthetic voice, and combination effect of female speech and synthetic voice all have significant effect on CAT test results in the presence of white noise. All tests used significance levels $\alpha = 0.5$.

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I gratefully acknowledge the valuable guidance and insight provided by my advisor and mentor, Dr. Misty Blue, in the successful completion of this research. Not only did she provide me with valuable insights in every step of the process, but also built my strength and self belief when the going got tough. Working under her capability was the best learning experience which taught me to view research in an elegant way. Numerous discussions with Dr. Blue have helped me to solve the research problem which played a vital role through this journey.

I would like to thank my committee members Dr. Yan Liu and Dr. Blair Rowley for being supportive and giving suggestions to improve the research work. Special thanks to Meg Wiltshire and Dr. Blue for providing me the recorded test samples for thesis without which this mile stone could not be achieved.

Working as a part of Biomedical Industrial and Human Factor Engineering Department at Wright State University was a wonderful experience. I am highly appreciative of all the support and help provided by the staff and students of the Biomedical Industrial and Human Factor Engineering Department.

I express my heartfelt appreciation to Institutional Review Board for providing required human subject protection training to begin testing the project on human subjects.

I appreciate the student participants for spending their valuable time to test the project. I would take this opportunity to thank my friends Rikki, Kutbi, Himanshu, Vinit, Pratik, Ruchit, Mihir, Bhavik, Darshan, Shital, Sakina, and Sachi for providing me a helping hand whenever needed.

Lastly, but most importantly I would express my love and gratitude towards my parents, Bipinchandra Soni and Urmila Soni, my brother Jignesh, sister-in-law Nilam and my nephew Yash, whose eternal support and motivation helped me smile through the thick and thin of this project.

1

Introduction

1.1 Background

Communication is vital in performing everyday activities. It is an exchange of information between a sender and a receiver via some medium. Speech communication is one of the most common forms of communication today.

Speech communication is a natural method to send and receive information and typically is done via some type of communication system (for example a telephone or PC call). When this is the case, for the communication to be effective, it must be intelligible whether it is produced by a machine or a human. Speech Intelligibility (SI) is defined as the degree to which speech can be understood and is measured subjectively with SI tests (Kent, 1992). SI is considered to be perfect when the listener is able to understand all the words intended to be communicated by the speaker. Speech is a vital part of military communication; it provides hands and eyes free operations. Speech is helpful in various military applications such as command, control, communication, computer, intelligence, and training. It is also used for specific applications for system control, workload reduction, training acceleration, and increment in situation awareness (Chen, 2005).

Effective and efficient speech communication is one of the leading factors for success of battlefield operations. Battleground speech communication is very critical for military operations (Gripper et al., 2007). There are many different noises that may

be present on the battlefield including noise from aircraft, small-arms and weaponry, vehicle noise, and sonic booms. Any of these if not properly managed can degrade progress. Also speech communication through radio transmission system is hindered by outside vehicle noise interference and other equipment noise. In military operations, unintelligible speech communication can be even more harmful than inaudible signals. For example, if a military person hears a signal, misinterprets it and takes action according to the random guess he made from the signal, it may result in a capture or even fatality (Gripper, 2006). In order to deal with critical battlefield situations such as noise, vibration, stress, poor radio channels, and high workload, speech communication systems must be tested for robustness prior to being used in the field. This is done primarily via speech intelligibility testing. Recently, the Auditory Research Team of the United States Army Research Laboratory, Human Research and Engineering Directorate (ARL-HRED), have made attempts to determine efficient speech communication methodologies for assessing speech intelligibility under realistic battlefield conditions.

Large amounts of the languages used in military speech communication are based on the phonetic alphabet and number codes. These types of phrases and codes are very similar across all military specialties including infantry, armor, medical personnel, and other military operatives (Rao & Letowski, 2006). In the past, the Modified Rhyme Test (MRT) has been widely used to test military communication systems but the language of these speech materials have been said to have little validity in military settings (Blue et al., 2004). In response to the criticisms, the Callsign Acquisition Test (CAT) was developed by the United States Army Research Laboratory especially for military use. It consists of 126 items or callsigns (Letowski, 2001; Blue et al., 2004). Each callsign is made up of a two-syllable word selected from the military alphabet (Alpha, Bravo, Charlie, etc.) followed by a one-syllable number (1 to 8, excluding 7). The CAT is a more appropriate intelligibility test for military applications because the callsign phrases are widely used in the military. Such test material combination is a good compromise between: (1) simplicity and poor predictive value of mono-

syllabic signals, and (2) the complexity and memory load of nonsense sentences and long number sequences. After initial positive findings and the "real-world"nature of the test, the CAT has been successfully applied in testing the effectiveness of various communication devices such as those that can be built into a soldier's helmet or worn over the ear (Henry & Mermagen, 2004). However, CAT still requires several steps of laboratory testing and field studies for validation before the final release of CAT may take place. One of these steps is evaluating the effects of voice on the CAT materials.

1.2 Speech Generation And Gender Differences

In the process of human voice generation, lung pressure generates a steady flow of air through the trachea, larynx, and pharynx. Fluctuations in air pressure created by the vibrations in the vocal folds are responsible for the generation of sound waves. Resonances in the vocal tract modify these waves according to the position and shape of the lips, jaw, tongue, soft palate, and other speech organs. Finally sound waves are radiated into environment through mouth and nose openings.

Most of the speech related research uses male voice as experimental samples. Past records show that in majority of military organizations, the greater part of the workforce consisted of men (Silverstein, 1953). As a result, primary users for communication mechanisms used in the battlefield were males. However, as per current scenario regarding the distribution of male and female population in the military, communication technology developers must be specific that their designs are able to enhance the performance of both genders. Amongst global characteristics, the effect of gender has been most prominent for the processing of speech signals (Markham & Hazan, 2004).

The most apparent dissimilarity between male and female voices is fundamental frequency (F0) or pitch. According to the Source-Filter Theory of Speech Production (Evans et al., 2008), formant frequency and fundamental frequency (F0) are two independent acoustic components of human voice. These components are determined by the size and the shape of the vocal tract and the vibration of the vocal folds

respectively. At puberty, several changes happen within a male human body including a permanent shift of fundamental frequency to a lower level due to changes in the larynx (Jenkins, 1998). Also the lower formant frequencies and less formant dispersion are generated by secondary fall of the larynx (Fitch & Giedd, 1999). This physical change within an adult male gives him a deeper and more imposing voice than an adult female. According to human morphology, a significant difference between male and female is the length of vocal tract and relative proportions of oral and pharyngeal cavities (Fitch & Giedd, 1999). An average speaking fundamental frequency for men typically ranges from 100 Hz to 146 Hz; whereas, that for women usually ranges from 188 Hz to 221 Hz (Gelfer et al., 2005). Few acoustic characteristics of female speech are lower in power, higher in frequency, and appear more susceptible than male speech in terms of getting masked by some of these military noises. These pitch levels help listeners in correct identifications of the speaker's gender.

Speech based systems offer an easy and natural way to communicate with computers. Computer-synthesized speech and a recorded human speech are the two basic alternatives considered when implementing technological speech output. Text-to-Speech (TTS) systems convert text input into real time speech output in variety of languages. AT&T Next-Gen, DECtalk, Prose-2000, and Infovox are few examples of TTS systems. TTS quality is characterized by two means; namely intelligibility of the produced speech and naturalness of the spoken synthetic speech. Formal tests with modern TTS systems showed that TTS word intelligibility is approaching that of naturally spoken speech (Schroeter, 2001).

A block diagram of a common concatenated TTS system is shown in figure 1.1. The first block is the text analysis module which consists of a series of modules with separate, but in many cases intertwined functions. Basic function of first module is to convert ASCII message text into a series of phonetic symbols and prosody (fundamental frequency, duration, and amplitude) objects. The second block in figure 1.1 is responsible for the advance towards much more natural sounding speech synthesis. It assembles the units according to the list of objects set by the forepart. Finally se-

lected units are fed into backend synthesizer which produces synthetic speech signals for listeners.

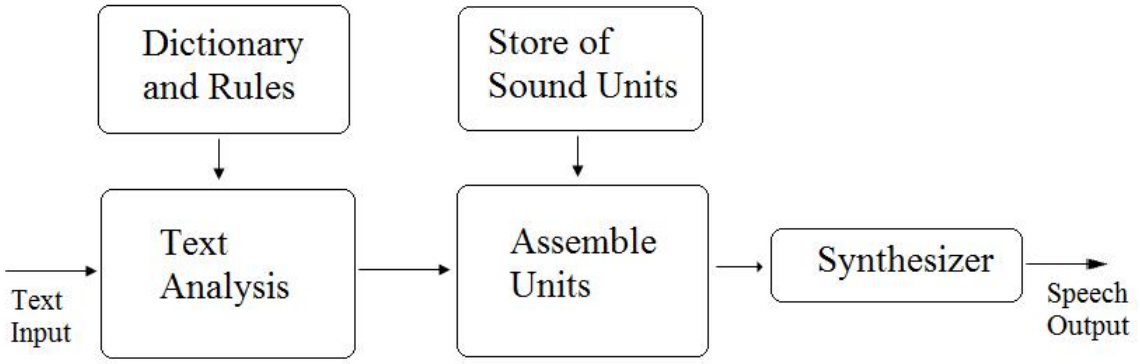

Figure 1.1: Block diagram of TTS synthesis system (Schroeter, 2001)

In this speech synthesis method actual segments of recorded speech are used which are stored in a voice database, either as "waveforms"(uncoded), or encoded by a suitable speech coding method. This speech synthesis then string together selected speech segments, and, after optional decoding, outputs the resulting speech signal. As this method uses speech segments form recorded speech, it contains the highest potential for sounding "natural".

Application of TTS systems has become widespread in areas such as telecommunications, information services, and disability services. Nowadays synthetic speech based voice response system has become widespread in consumer products, industrial and military applications, and aid for the handicapped. Due to increased use of Text-to-Speech (TTS), there is a need for assessment procedures that can reliably differentiate the usefulness of the systems in a variety of contexts (Van Santen et al., 1998). Speech intelligibility is one of the important factors to be considered while selecting specific application based voice response system. Speech coding systems generate a relatively limited vocabulary of utterances using a fixed set of parameters; whereas synthetic speech can automatically convert unrestricted text into speech. Significant researches on synthetic speech are needed to make them more intelligible and very natural (Klatt, 1983). There are no existing methods for automating the

assessment of synthetic speech quality; therefore, ultimate test involves assessment of human perceptual responses (Pisoni et al., 1985). To incorporate synthetic speech intelligibility test results for real-life situations, the present study assessed synthetic CAT speech in white noise.

Some findings illustrated that listeners were more comfortable with male synthetic voice than the female synthetic voice (Nass et al., 2003). Though male speech is considered more intelligible than female speech, it has been proven that the effect of signal quality and speech gender is not consistent for all TTS systems (Stevens et al., 2005). Few studies related to gender effect demonstrated that male and female participants are more positively oriented in terms of trust towards their own gender voices, even for the synthetic speech (Lee et al., 2000; Eagly, 1983). According to a general study, compare to male listeners female listeners are more sensitive to the difference between recorded speech and synthetic speech (Nass et al., 2003). Listeners seem more willing to reveal to human speech that provide the illusion of more recognizable social interaction than robotic speech which may emphasize on its non-human origin (Nass et al., 2001).

Military TTS systems support interactive training and learning environments by using computer supported learning systems. Some TTS systems help train personnel with minimum or no instructor participation (Chen 2005). Speech technologies offer training at reduced cost and timing. According to Smither (1993), recall of eight-digit numbers was considerably better when the numbers were presented by a male human speech than by a synthesized male speech, even when recorded speech was as intelligible as synthesized speech. Some of the performance related limitations with TTS system can be overcome by listener's adequate training, experience, and practice. Learning of structural characteristics in TTS system make listener better perform in intelligibility tests (Pisoni et al., 1985).

1.3 Research Objectives And Scope

The number of females in the armed forces has increased with the formation of all-volunteer forces. A significant number of females are currently employed for control tower communication at civilian and military airfields (Silverstein, 1953). The increased number of females using both civilian and military communication systems lends itself to the need to examine the predictive power of the CAT materials to determine its applicability for female speakers. The primary objective of this study was to determine the effect of voice (male and female) and speech production (natural and synthetic) on the CAT's predictive power of SI. The research presented here will also provide valuable data to aid in the standardization and validation process of the CAT test material.

1.3.1 Key Research Questions

This research study emphasizes on the effects of gender using white noise at two different signal-to-noise ratios (SNRs), for both male and female voice, in both natural speech and synthesized speech. SNR is defined as the difference between the output level of the background noise and the speech. In an attempt to complete this research work towards validation of the CAT test materials, answers to the following question will be determined:

1. Are there any significant differences in speech intelligibility scores between the male recorded version of CAT and the female recorded version of CAT?

2. Are there any significant differences in speech intelligibility scores between the synthetic voice and the human recorded voice?

1.4 Importance Of The Research Questions

The following explanations are provided to gain further understanding of above stated research question.

1.4.1 Impact Of Human Recorded Voice (Male Versus Female) On CAT Intelligibility

This research question compares the effect of the speaker's voice on CAT SI scores. It has been determined that in normal living conditions the effect of different genders on the speech intelligibility is similar (Ellis et al., 1996; Ellis et al., 2002). However, in noisy environments, research indicates that male voice is more intelligible than female voice (Nixon et al., 1998). One probable reason for these findings is that the fundamental frequency of the female voice may be closer to the ambient noise. Thus the listener will receive the masking portion of the signal (Nixon et al., 1998). With growing numbers of females enlisting in the military, the communication systems need to be tested for robustness with both male and female voice. This can be done in two ways: (1) with materials that are proven to provide intelligibility predictions that are voice independent, or (2) test in both male and female voice. Currently most speech intelligibility tests are available in both male and female voice and testing is required using both sets of materials. This can be time consuming as well as costly. If it can be shown that the CAT materials are voice independent (there are no differences in SI scores for male and female voice), this could be a major advantage over existing tests.

For this research study, three possible outcomes are: (1) intelligibility scores are significantly higher for male voice, (2) intelligibility scores are significantly higher for female voice, or (3) there are no differences between the intelligibility scores for male and female voice. If the results do not show any significant effect of gender on CAT test results, then, it would prove the CAT to be voice independent and superior to other SI tests in its ability to predict SI in the contexts where gender of the talker is an issue.

1.4.2 Impact Of Speech Synthesis On CAT Intelligibility

The purpose of this question is to determine the effect of speech synthesis on CAT SI scores. Also the effect of gender on the SI scores for CAT when the speech is

synthetic is examined. Intelligibility and naturalness are two important qualities of any synthetic system. Some findings on synthetic speech systems demonstrated that as synthesized speech systems lacks the quality of natural human speech; naturalness scores even for their best systems are in the poor-to-fair range (Nass & Lee, 2001; Kamm et al., 1997). Also "Synthetic speech systems tend to have inexplicable pauses, misplaced accents and word emphases, discontinuities between phonemes and syllables, and inconsistent prosody". (Nass & Lee, 2001). In contrast, several research studies on synthetic speech proved that synthetic speech obtains similar social responses as recorded speech (Lee et al., 2000). Many industries are turning to synthetic speech for communication systems. The predictive capabilities of CAT in synthetic speech in both male and female voice are very useful as this trend continues.

1.5 Organization Of Thesis

This research provides an overview of speech intelligibility and testing (Chapter 2), experimentation, methodology, and results (Chapter 3) for different experiments that were designed to answer the above stated research questions. Finally, a general discussion and conclusion (Chapter 4) provides summary of the research.

2

Speech Intelligibility And Testing

2.1 Speech Intelligibility

As there are many sources of potential error, the process of communication is very complex whether between two people or a person and a technology. Common causes of error for human-to-human speech communication would include the speaker's accent or tone. For human-to-machine (and vice versa) communication, potential sources of error would include a poorly designed interface or faulty equipment (Gripper, 2006). Sometimes, due to lack of intelligibility, the meaning of information is lost when it is transmitted via a communication system. According to social psychologists, during the transmission of messages from sender to receiver, usually 40-60% meanings are lost (Wertheim, 2008). For the selection, use, and development of speech communication devices the evaluation of speech intelligibility is a significant phase and it should be repeated at various stages throughout speech technology development (Sydral et al., 1994).

In general terms, the science of error analysis of human speech is known as the science of intelligibility (Allen, 2005). Speech intelligibility may be defined as the match between listener's responses to the intention of the speaker's speech passed through a transmission system. Speech with higher intelligibility would be easier to recognize. Lower intelligibility of speech leads to misinterpretation, confusion, and missing of words (Sydral et al., 1994).

During speech intelligibility testing, the listener will listen to a series of test items and make their best attempt to identify the words or sentences they heard. To assess speech communication systems, there are two categories of measures for speech intelligibility: (1) perceptual measure use human listeners to measure intelligibility of communication systems, and (2) technical measure that use a standardized signal broadcast over the communication system to predict its intelligibility based on the physical parameters of the transmission channel (Meyer Sound Labs, 2008). Speech intelligibility tests such as the CAT fall in the perceptual measures category. The following sections give detailed explanation of perceptual measures.

2.2 Perceptual Measures Of Speech Intelligibility

This speech intelligibility measure uses humans to evaluate voice communication systems. In this measure, normal hearing listeners are asked to identify the words or phrases which are presented over the system under testing. For this assessment transmitted speech materials are selected from standardized materials which are specifically designed to assess particular characteristics of speech communication. These materials are typically spoken and pre-recorded by trained speakers to eliminate the errors generated by variation in the speaker's speech volume. Depending upon the evaluation purpose, word samples are transmitted via headset, bone conductor, or loudspeaker system. To compare test results with real life speech communication situation, test may be executed in many different acoustic environments. As long as human subjects are used for experiments, the experimenter should consider experiment related psychological issues, for example, effects of training, memorization, repetition, and wordlist presentation order (e.g., ascending order, descending order or random order). For these experiments larger subject groups are selected to reduce the error margins. Final results are calculated in terms of percentage correction ratio.

Some disadvantages of perceptual metrics are human errors, long preparation and testing time, higher expenditure for the test, and highly situation specific characteristics. Speech intelligibility rating test (SIR) and performance intensity (PI) function

are two common perceptual metrics for speech intelligibility. The SIR test was developed for the evaluation of hearing aid settings. In this test listeners listen and rate the intelligibility of test material according to their understanding while wearing the hearing aid under test. The PI function describes how intensity of the test material affects the speech recognition performance of the listener. It is the end result graph where intensity is plotted on x-axis and average percentage correct scores on y-axis.

2.3 Speech Audiometry

Nowadays speech and hearing are vital factors for education and communication (MacFarlan, 1945). Good hearing is required for efficient interpretation of speech. The process of hearing measurement is commonly termed as "Audiometry". It is usually performed by monitoring, recording, and analyzing listener's responses to controlled acoustic stimuli in clinical settings, in which standardized language sample is presented through calibrated system (audiometer)(Carhart, 1946). Pure tone audiometry and speech audiometry are most commonly used audiometric screenings in clinical settings. Pure tone audiometry uses tonal (tones of only one frequency) signals varying from lower to higher pitch sounds to assess hearing capabilities in humans. Speech audiometry assesses hearing capability as well as comprehension as it uses speech stimuli to assess the functionality of the human hearing sense. Speech audiometry techniques are also used in communication system testing to assess important information regarding the integrity of audiometric systems (Konkle & Rintelmann, 1983).

Pure tone testing helps in finding severity and configuration of hearing loss and indicates the supposed anatomical site of lesion. The most frequent complaints by the persons with hearing loss are concerned with the understanding of speech. It is obvious that, as pure tone audiometry test provides data for basic hearing thresholds, but it gives little to no information about the extent to what is being understood of what is perceived. Alternative stimuli must be used to assess this communicative skill. Speech audiometry additionally gives information about the sensitivity to speech

material, and speech understanding at supra-threshold level.

In broad terms, speech audiometry is a procedure concerned with determining whether or not a person is able to understand the word-lists presented to them. The Speech Awareness Threshold (SAT) also known as Speech Detection Threshold (SDT) is the test to estimate the lowest speech hearing level at which an individual can just detect the presence of a speech material at least half of the time. The participants in this test are required to show awareness of the sound presence but not required to identify the material as speech. Speech reception threshold (SRT) is a one of the basic measurement in speech audiometry. SRT determines the lowest intensity level at which 50% of common two-syllable words (e.g., cowboy, toothbrush, playground etc.) can correctly identify by an individual.

A speech discrimination test (also known as Word recognition test) determines person's ability to understand speech in terms of percentage correct score by presenting louder speech which is well above individual's threshold level. Depending upon the cause of the hearing loss, there should be a correlation between the word recognition score (WRS), and the type and degree of hearing loss (Hain, 2003). The WRS can help predicting the effectiveness of the hearing aid. If there is an increase in the WRS with amplification, it is suggested that hearing aid may be useful for hearing.

2.3.1 Testing Materials

Basic procedures for speech intelligibility testing require listeners to respond to a standardized set of materials and record what they are able to understand of what they hear. Primarily depending upon the purpose of the evaluation, speech material will be selected as stimuli for hearing measurement. Whereas pure tone audiogram is unable to derive information about communicative abilities, it is impractical to anticipate a given sample of speech to be the representative of all type of verbal communication. Thus, depending upon the nature and the purpose of the test various types of speech materials are used in speech audiometry. The following section provides a basic description about commonly used forms of speech testing materials.

The speech testing words and sentences must be phonemically balanced (PB) for a particular language. In the Phonetically Balanced word list, to estimate the relative frequency of phoneme occurrence in each language, the monosyllabic test words are selected (Logan et al. 1989). Different kinds of word lists and sentences lists are available for speech intelligibility assessment. To assess the listener's ability to identify phonemic unit of speech in audiometry, syllables are used. They are generally in sequence of vowel-consonant (VC), consonant-vowel (CV), or consonant-vowel-consonant (CVC). Syllables used for testing are of two types, monosyllabic words or nonsense stimuli. Meaningful monosyllabic words are the most popular and widely used hearing assessment speech material. Some advantages associated with monosyllabic words are convenience for constructing audiologic tests, and ease of developing word lists with several equivalent alternative words (Konkle & Rintelmann, 1983). Nonsense stimuli help in avoiding influence of listener's vocabulary on correct recognition of phonemes. Digits are also used for hearing evaluation. They play important role in speech threshold testing.

Spondaic words or spondees are two syllable words with equal stress on each syllable. Some examples of spondaic words are cowboy, toothbrush, and playground. Primarily to determine various types of thresholds spondees are used. Spondaic words help in determining Speech Reception Threshold (SRT) by calculating the lowest intensity level at which 50% of common two-syllable words can be correctly identified by an individual. A characteristic like homogeneity for intelligibility makes them popular and appropriate for threshold measurement in speech audiometric assessment (Eagan, 1948).

Speech audiometric assessment materials such as syllables and words have been criticized for their failure in representing stimuli that encountered in real life communicative conditions. Sentence materials are usually seen as having greater representation of everyday communication. The Speech Perception in Noise (SPIN) Test, the Synthetic-Sentence Identification (SSI) Test, and the Everyday Sentence Test are some general sentence identification tests.

2.3.2 Speech Intelligibility Tests

There are different kinds of word lists and sentence lists that are available for speech intelligibility assessment. The Modified Rhyme Test (MRT) is currently one of the most commonly used word lists for military as well as for civilian applications. The CAT is under evaluation for future standardization. This section provides basic information about Modified Rhyme Test (MRT) and Callsign Acquisition Test (CAT).

Modified Rhyme Test (MRT) is a modified version of the original rhyme test which uses lists of rhyming monosyllabic words (House et al., 1965). MRT is most commonly used speech intelligibility test for military communication system (House et al., 1965). This test contains a total of 300 words divided as 50 sets of 6 words. The MRT uses rhyming words to test for initial and final consonant comprehension. Examples of those sets which tests for final and initial consonants are given in table 2.1 and table 2.2 respectively. In testing with the MRT wordlist, listeners are given the set of rhyming words (6 words). One word from the set is presented to listener via spoken voice and listener has to identify or make a guess about the word that they heard from the particular set given to them. For MRT testing ceiling effects are introduced due to high predictability of the responses (Handley, 2008). The vocabularies used in the MRT are not characteristic of words typically used in military settings. For this reason, it has been criticized as having poor validity in military settings.

Word 1	Word 2	Word 3	Word 4	Word 5	Word 6
bus	but	bug	buff	bun	buck

Table 2.1: Example of MRT words used in testing for final consonant apprehension

Word 1	Word 2	Word 3	Word 4	Word 5	Word 6
led	shed	red	fed	bed	wed

Table 2.2: Example of MRT words used in testing for initial consonant apprehension

The CAT test was developed with the shortcomings of MRT in mind by the United States Army Research Laboratory (Letowski, 2001; Blue et al., 2004). To

reduce between-letter confusability and improve performance for both human-human and computer-human communication phonetic alphabets are generally used (Gripper, 2006). Phonetic alphabets are used greatly in vocabulary of military settings. Listeners typically find the testing material more intelligible if they are more familiar with the testing material. For military communication system testing, this phenomenon means a lot. Accurate speech intelligibility scores can and should be obtained by testing end users as subjects (military personnel). More accurate intelligibility ratings will be achieved with familiar speech material for military communication testing.

The full version of CAT consists of 126 items or callsigns. Each callsign is made up of a two-syllable word selected from the 18-item military alphabet (Alpha-Zulu) followed by a one-syllable number (all numbers from 1 to 8, excluding 7). All CAT alphabetic codes and digits are shown in Table 2.3. The CAT is a more appropriate intelligibility test for military applications because call sign phrases are widely used in the military. Such test material combination is a good compromise between (1) simplicity and poor predictive value of monosyllabic signals, and (2) the complexity and memory load of nonsense sentences and long number sequences. Digits are used for their high intelligibility, familiarity, audibility homogeneity, and easy production as spondee.

Alphabet Codes			**Digits**
A(Alpha)	H(Hotel)	T(Tango)	1
B(Bravo)	K(Kilo)	V(Victor)	2
C(Charlie)	L(Lima)	W(Whiskey)	3
D(Delta)	O(Oscar)	X(X-Ray)	4
E(Echo)	P(Papa)	Y(Yankee)	5
F(Foxtrot)	Q(Quebec)	Z(Zulu)	6
			8

Table 2.3: Alphabetic codes and digits used for Callsign Acquisition Test (CAT)

3

Methodology And Results

3.1 Background

As there are always many different types of background noise present in a military environment (e.g., small-arms and other blast noise, artillery noise, vehicle noise, sonic booms etc.), we can't even imagine any real-time military situation without noise. If we consider testing research samples without noise, our research conclusion will not provide a practical judgment about the intelligibility of speech in real life situation. For the present study, to evaluate CAT files in real life situation, white noise was selected as background noise.

White noise is a noise that is combination of sound from all different frequencies in equal amount. The White Noise, by analogy with the white light, which contain power spectral density spread over the visible band. As white noise consists all frequencies in audible range, in which all components are at same amplitude, it is commonly used to mask other speech signals.

3.2 Objectives

There are obvious differences between male and female voice; the most apparent being fundamental frequency (F0) or pitch. The female voice is typically higher in pitch. This has lead to the need for SI tests to be recorded in both voices. In the past all SI tests have been recorded in both male and female voice because of differences in

the predictive power of the materials between the two voices. Thus far, there have been no data collected comparing the predictive power of the CAT materials in the two different voices. Predictive power of the testing material refers to its capability to produce testable predictions. This research is necessary for two related reasons: (1) to complete the standardization process as recommended by the ANSI standards, and (2) to determine if the CAT test needs to be recorded in both male and female voice. After approval from the institutional review board of Wright State University, the data was collected as outlined in the following paragraphs.

3.3 Methodology Of Data Collection

3.3.1 Participants

A group of 21 listeners between the ages of 18 and 26 participated in the study. Of the participants, 14 were male and 7 were female. All participants were required to have pure-tone hearing thresholds better than or equal to 20 dB HL at audiometric frequencies from 250 Hz through 8000 Hz (ANSI S3.6-1996). Audiometric screening was performed by one of the investigators after the consent form was signed. The hearing test involved standardized clinical equipment and procedure. All participants were recruited from the student population of Wright State University.

3.3.2 Instrumentation

Instrumentation for the study included: (1) Personal Computer with CD ROM drive, (2) a two channel clinical audiometer Equinox, (3) CD ROM with CAT test materials recorded in four different voices (human recorded male and female and synthetically produced in male and female), (4) a white noise file presented at 65dB, (5) Crown D-75A audio amplifier connected in series between the computer and the speakers, (6) Extech Instruments 407735 Sound Level Meter, (7) an acoustically treated sound booth, and (8) a pair of testing loudspeakers.

3.3.3 Sound Files

Four different sound files used for this experiment were in male natural voice, male synthetic voice, female natural voice, and female synthetic voice. Synthetic voiced files were generated through AT&T Next-Generation TTS (Text-to-Speech) system. The natural speech voices were recorded and normalized to be within 1dB. A white noise file and the four different CAT sound files were edited with Sony Sound Forge 9 software.

Speech-to-Noise Ratio (SNR) is the difference between intensity of the speech signal and the background noise. It is measured in terms of decibels (dB). The SNR's used for this study were -12dB and -15dB. In order to get SNR at -12 dB and -15 dB, white noise file was presented at 65 dB and speech files were presented at 53 dB and 50 dB respectively. All these sounds were at or below the levels for normal conversation, thus no danger was presented to the participants hearing. All different combinations of noise and speech level used in this experiment are shown in Table 3.1.

Test Condition	**Speech Type**	**Speech Level**	**Noise Level**	**SNR**
T1	Male Natural	50 dBA	65 dBA	-15 dB
T2	Male Natural	53 dBA	65 dBA	-12 dB
T3	Male Synthetic	50 dBA	65 dBA	-15 dB
T4	Male Synthetic	53 dBA	65 dBA	-12 dB
T5	Female Natural	50 dBA	65 dBA	-15 dB
T6	Female Natural	53 dBA	65 dBA	-12 dB
T7	Female Synthetic	50 dBA	65 dBA	-15 dB
T8	Female Synthetic	53 dBA	65 dBA	-12 dB

Table 3.1: Noise and speech levels for each speech-to-noise ratio for male and female speakers

3.3.4 Procedure

Each of the 21 listeners participated in a single listening session that consisted of a hearing screening, a brief training, and eight SI tests. Each participant was seated inside an acoustically treated booth in front of a computer monitor and keyboard.

The computer keyboard was used to record responses. The participants were asked to listen to the series of CAT test items and identify them by pressing appropriate keys on the keyboard. If the listeners were unsure of what they heard, they were instructed to make their best guess. For example, if they heard Bravo 6, the correct key response would be "B, 6, Enter". Figure 3.1 shows computer screenshot of CAT software.

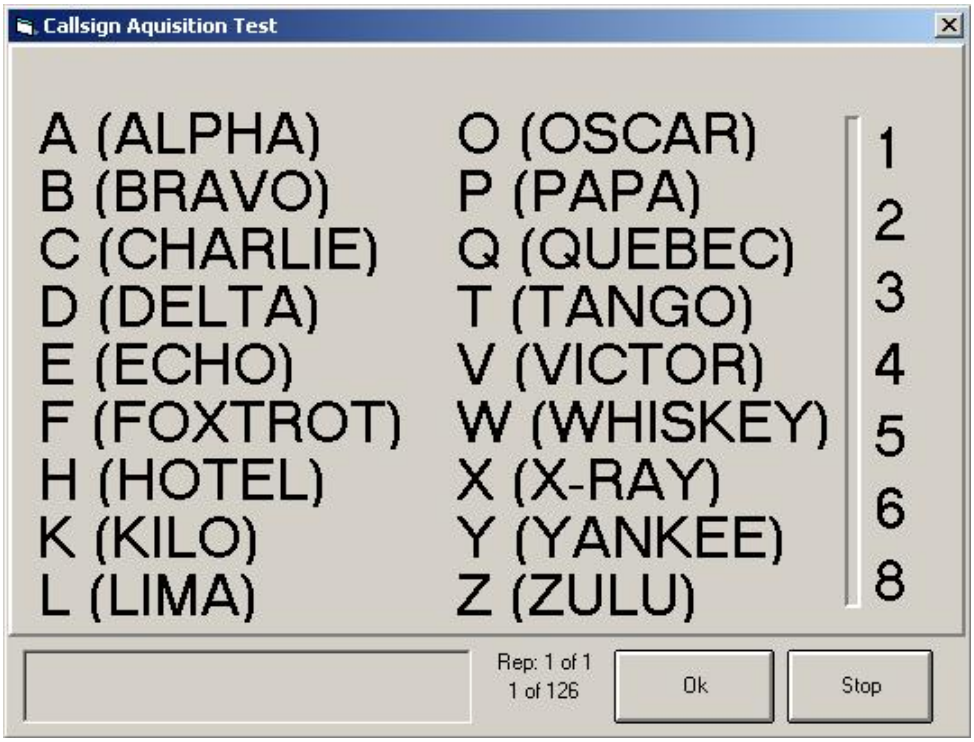

Figure 3.1: Computer screenshot of CAT software

All sounds were presented via loudspeakers places at 135° and 225°. The four lists were (1. male natural recorded voice, 2. female natural recorded voice, 3. male synthetic voice, and 4. female synthetic voice) presented (in a random order). Each of the lists were presented at the two different SNRs (-12dB and -15dB), thus each participant listened to eight CAT presentations.

Each SI test usually takes about 7 minutes. The entire session was lasted about 1 hour and 40 minutes. This included 20 minutes for a hearing screening, 10 minutes for

consent and briefing, 8 SI tests, 7 minutes each (56 minutes), two 5 minutes breaks, and 5 minutes to debrief.

3.3.5 Experimental Design

The independent variables for this study were the SNR (two levels), voice of CAT test materials (Male and Female), and CAT test material type (Natural and Synthetic speech). The dependent variable was the listeners scores expressed as a percentage of correct responses. To avoid any order or learning effects, presentation of individual systems was randomized as shown in Table 3.2 (Byers, 1991). This 8×8 Latin Square was repeated for every block of 8 listeners. Thus, listeners 1, 9 and 17 listened to the same order of CAT test items.

	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**L8**
Condition 1	M S 12	F S 12	F S 15	M S 15	M N 15	F N 12	F N 15	M N 12
Condition 2	F N 15	M S 15	F S 12	M S 12	M N 12	M N 15	F N 12	F S 15
Condition 3	F S 15	M N 15	F N 12	M N 12	M S 12	M S 15	F S 12	F N 15
Condition 4	M N 15	F N 15	M S 12	F N 12	F S 12	F S 15	M N 12	M S 15
Condition 5	F N 12	M S 12	M S 15	F N 15	F S 15	M N 12	M N 15	F S 12
Condition 6	M N 12	F N 12	F N 15	M N 15	M S 15	F S 12	F S 15	M S 12
Condition 7	M S 15	F S 15	M N 12	F S 12	F N 12	F N 15	M S 12	M N 15
Condition 8	F S 12	M N 12	M N 15	F S 15	F N 15	M S 12	M S 15	F N 12

Table 3.2: 8×8 Latin square for counterbalancing experimental conditions (M=Male, F=Female, N=Natural, S=Synthetic, 12=-12dB SNR, and 15=-15dB SNR)

3.4 Hypotheses

A hypothesis addressed with this experiment is stated below:

- Hypothesis: There are no statistically significant effects of individual and combined independent factors (Gender, SNR, and Type) on the intelligibility scores for CAT in white noise.

3.5 Data And Data Analysis

Factorial analysis was performed to analyze the experiment data. All raw data collected from the experiments are summarized in Table 3.3.

Gender	Type	SNR(dB)	Mean % Correct Score	Std. Deviation % Correct
Male	Natural	-12	94.05	5.61
		-15	87.90	7.48
	Synthetic	-12	92.48	4.85
		-15	81.67	10.12
Female	Natural	-12	93.10	5.87
		-15	84.90	6.14
	Synthetic	-12	79.71	7.65
		-15	71.00	7.79

Table 3.3: Speech intelligibility scores of experiment

To form a performance intensity (PI) function, mean percentage correction scores are plotted against SNR. Figure 3.2 shows PI function for male natural and female natural speech, in which mean percentage correction scores for male natural and female natural speech are plotted against -12 dB and -15 dB SNR. As shown in graph below, male natural speech provides higher intelligibility results than female natural speech. Also Nonparallel PI functions of male and female natural recorded voice demonstrate its dependency of noise level. Figure 3.3 shows PI function for male synthetic speech and female synthetic speech. It illustrates the same intelligibility results as natural recorded voice. Also parallel PI functions of male and female synthetic voice display its independency of noise level. Statistical proofs of these results are explained in next section.

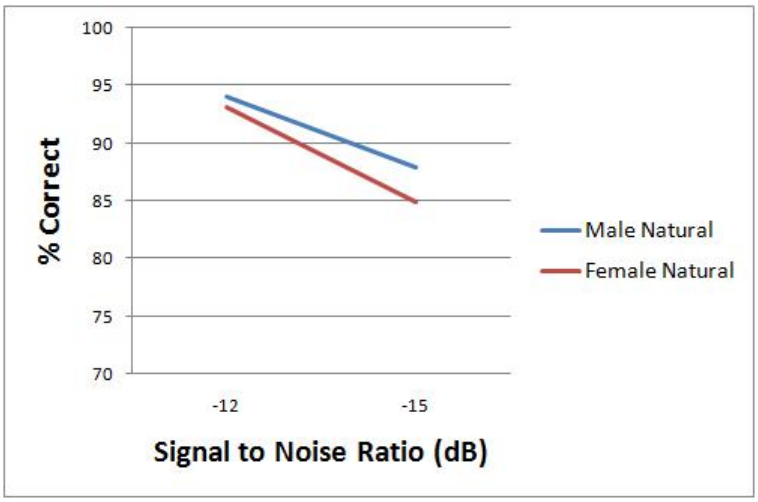

Figure 3.2: PI function for male natural and female natural speech at -12 dB and -15 dB SNR

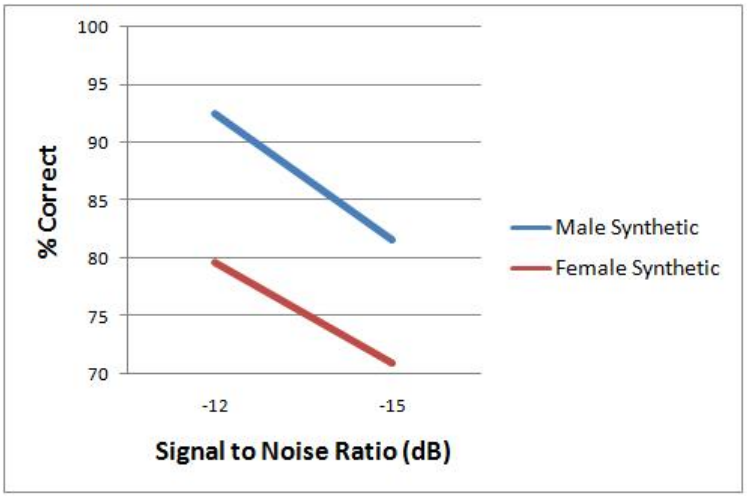

Figure 3.3: PI function for male synthetic and female synthetic speech at -12 dB and -15 dB SNR

As revealed in figure below, human natural speech has higher mean intelligibility scores than synthetic speech.

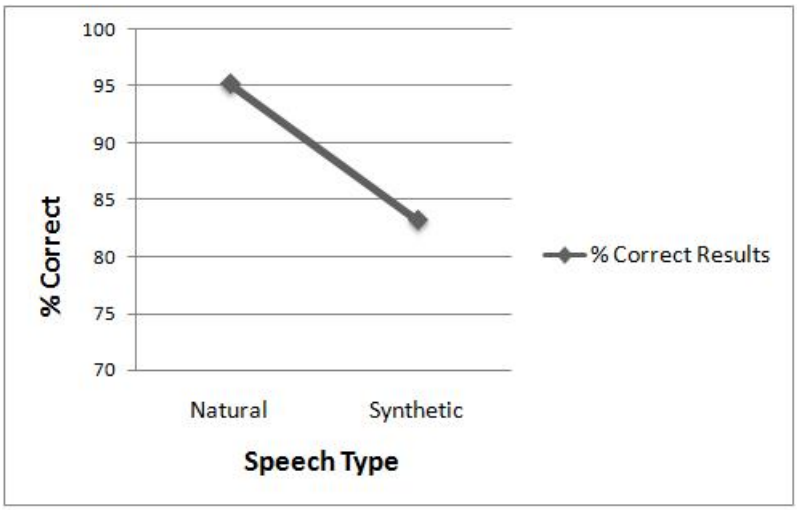

Figure 3.4: Mean percentage correct scores for natural speech and synthetic speech

3.5.1 Statistical Calculations

A factorial analysis was performed to demonstrate the significance of each factor. Before analysis all percentage correction score were stabilized with arcsine transformation to accommodate ceiling effect of both end of the evaluation scale. To convert the scores into rationalized arcsine units (rau), Studebaker's (1985) "rationalized"arcsine transform was used. The arcsine transformation help analyst when working with proportions and percentages. Rationalized arcsine units (rau) conversion is advantageous over other arcsine transforms, as they are closer to the original percentage scores which help in comparison of rau data with original data. For this experiment value of α was set to 0.05.

A null hypothesis (H_0) and alternative hypothesis (H_1) addressed with this experiments are stated below:

- H_0 = There are no main and interaction effects of independent factors (Gender, SNR, and Type) on the intelligibility scores for CAT in white noise.

- H_1 = At least one main or interaction effect of independent factors significantly affects the intelligibility scores for CAT in white noise.

As there were more than two factors (Gender, SNR, and Type) present at a same time full factorial analysis was performed. It evaluates main and combined effect of independent factors on test results. Main effect is an effect of a factor alone on a dependent variable whereas, interaction effect is due to combination of two or more factors present. A three-factor analysis of variance consists of seven significance tests: a test for each of the three main effects, a test for each of the three two-way interactions, and a test of the three-way interaction.

Before factorial analysis all independent factors data should be normalized into common level. For this experiment all independent factors have two levels such as male and female for gender, -12 dB and -15 dB for SNR, and natural and synthetic for type. All levels are converted in terms of (+1) and (-1) for this analysis. Table 3.4 indicates all independent factors, levels and according coding. Statistical calculations of factorial analysis are shown below.

Factors	**Levels**	**Coding**
Gender	(Male, Female)	(1,-1)
SNR	(-12dB, -15dB)	(1,-1)
Type	(Natural, Synthetic)	(1,-1)

Table 3.4: Independent factors and according level coding

Response % Correct

Summary of Fit

RSquare	0.789215
RSquare Adj	-0.76005
Root Mean Square Error	18.34706
Mean of Response	89.24875
Observations (or Sum Wgts)	168

Parameter Estimates

Term	Estimate	Prob>|t|
Intercept	89.24875	<.0001*
Gender[-1]	-4.592202	0.0041*
SNR[-1]	-6.243036	0.0003*
Gender[-1]*SNR[-1]	0.3243452	0.8211
Type[-1]	-6.012679	0.0004*
Gender[-1]*Type[-1]	-3.043393	0.0440*
SNR[-1]*Type[-1]	-0.10375	0.9423
Gender[-1]*SNR[-1]*Type[-1]	1.0891071	0.4506

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob > F
Gender	3542.84	3542.84	1	108.4908	<.0001*
SNR	6547.88	6547.88	1	196.5778	<.0001*
Gender*SNR	17.6736	17.6736	1	0.4201	0.5243
Type	6073.59	6073.59	1	84.8294	<.0001*
Gender*Type	1556.06	1556.06	1	23.0847	0.0001*
SNR*Type	1.80836	1.80836	1	0.0314	0.8611
Gender*SNR*Type	199.274	199.274	1	3.3895	0.0805
Gender*Subjects&Random	653.113	32.6556	20	0.6443	0.7827
SNR*Subjects&Random	666.187	33.3094	20	0.8156	0.6690
Gender*SNR*Subjects&Random	841.448	42.0724	20	0.7156	0.7695
Type*Subjects&Random	1431.95	71.5977	20	1.0820	0.4844
Gender*Type*Subjects&Random	1348.13	67.4065	20	1.1465	0.3814
SNR*Type*Subjects&Random	1151.14	57.5569	20	0.9790	0.5187
Gender*SNR*Type*Subjects&Random	1175.82	58.7911	20	0.1747	0.9999

Figure 3.5: Factorial analysis

Based on the F-ratio and p-values (< 0.05) of the effects we might conclude that female speech (Gender [-1]), -15 dB SNR (SNR [-1]), synthetic speech (Type [-1]), and interaction effect of female speech and synthetic speech (Gender [-1]*Type [-1]) have significant effect on CAT test results in white noise. From the PI function plot we can easily conclude that male speech has better intelligibility results than female speech and human natural speech reveals greater intelligibly score than synthetic speech.

4

Conclusion And Discussion

The primary purpose of this research was to aid in validation and standardization process of CAT material. In this experiment, we examined the effect of gender using white noise at 2 different signal-to-noise ratios (SNRs) for both male and female voice in both natural recorded speech and synthesized speech. Within the limitations of the current study, several conclusions can be made. The current study provides supporting evidence of a significance of all three independent factors: speakers' gender, noise level, and speech type on CAT material in presence of white noise. Speech intelligibility data collected for this experiment are expressed as percentage correct of item correctly perceived. For further statistical analysis data are stabilized with rationalized arcsine transformation. All research questions will be answered in next section.

4.1 Are there any significant differences in speech intelligibility scores between the male recorded version of CAT and the female recorded version of CAT?

The PI functions for human natural speech and synthetic speech answers this research question. In both PI functions, the male voice of the CAT material gave higher intelligibility scores than female voice material. Mean percentage correction scores and standard deviations (SD) for male and female CAT items were 89.02 (SD=8.65) and

82.18 (SD=10.55) respectively. In order to answer this research question statistically, a factorial analysis was performed. Based upon the statistical analysis, participants more precisely identified CAT material presented in male voice when compared to those presented in the female voice when white noise was presented. The lower fundamental frequency of the male voice may result in a slightly higher level of understanding than the higher fundamental frequency of the female voice in the current experiment due to the frequency composition of white noise and its ability to mask female voice more efficiently.

For this experiment, CAT material was evaluated at two different SNR levels -12 dB and -15 dB. As per statistical calculations, noise factor significantly affect the test results alone. Both PI functions and statistical analysis indicated that as the difference between the noise level and the speech level increased, the accuracy on the CAT material decreased overall. Mean percentage correction score and standard deviations (SD) for CAT material at -12 dB and -15 dB were 89.83 (SD=8.40) and 81.37 (SD=10.15) respectively.

4.2 Are there any significant differences in speech intelligibility scores between the synthetic voice and the human recorded voice?

The statistical analysis and mean percentage correction scores showed that synthetic speech does have significant effect on the intelligibility score of the CAT. Mean percentage correction scores and standard deviations (SD) for natural speech and synthetic speech are 89.98 (SD=7.26) and 81.21 (SD=10.87) respectively showed higher intelligibility of human natural speech. Even for the synthetic speech male voiced CAT material revealed higher percentage correction scores.

4.3 General Discussion

Speech technology has long research history in the military applications. Due to the critical nature of the tasks carried out by military operations, it is essential to make sure that new speech intelligibility test is able to perform effectively in a variety of military environments, such as various level of ambient noise. Also with increase in the level of gender diversity in military services, it is important to assess effectiveness of voice for both genders. Results of the current study confirm that speech intelligibility score for male voiced CAT material is better than that of female voice in presence of white noise whether it is formed by human natural voice or synthetic voice. Before the final release of CAT test, field validation under real-world military communication is desirable.

4.4 Future Research Recommendations

From this research on the perception of CAT material in presence of white noise, we have been able to specify the effect of gender and speech synthesis on speech intelligibility results. However, there is still more research to be done. Basic research is to evaluate human natural speech and synthetic speech into severe military environments, how perception is influenced by prior experience, and naturalness of SI testing material. Further test should be conducted using several male and female voices in a wide range of fundamental frequencies. The current study investigates intelligibility results in presence of white noise only, so further investigation with different noise condition is desirable. In short period of time advancement in speech synthesis technology is enhanced by the development in the microcomputer industry. Nowadays speech synthesis considered as one of the essential aspects of virtual reality. Synthetic speech is a future of speech system in future combat, so additional research is needed to recognize the role of practice and familiarity on synthetic speech testing material.

5

References

Allen, J.B. (2005). Articulation and Intelligibility. San Rafael, CA: Morgan and Claypool

Blue, M., Ntuen, C. & Letowski, T. (2004). Speech Intelligibility of the Callsign Acquisition Test in a Quiet Environment. International Journal of Occupational Safety and Ergonomics (JOSE), 10(2), 179-189.

Byers, J.A. (1991). BASIC algorithms for random sampling and treatment randomization. Computers in Biology and Medicine 21:69-77.

Carhart, R. (1946). Tests for the selection of hearing aids. Laryngoscope, 56:780-794.

Chen, F. (2005). Designing Human Interface in Speech Technology. Springer. p. 289-330.

Eagan, J. P. (1948). Articulation Testing Methods. Laryngoscope, 58, 955-991.

Eagly, A. H. (1983). Gender and social influence: Asocial psychological analysis. American Psychologist, 38, 971-981.

Ellis L. W., Spiegel B. & Benjamin B. (2002). Effects of speakers' augmented characteristics and listeners' sex on intelligibility and acceptability of synthesized speech, Perceptual and Motor Skills 94, 1081-1088.

Ellis L., Reynolds L., Fucci D. & Benjamin B. (1996). Effects of gender on listeners' judgments of speech intelligibility, Perceptual and Motor Skills 83, 771-775.

Evans S., Neave N., Wakelin D., & Hamilton C. (2008). The relationship between testosterone and vocal frequencies in human males. Physiology and Behavior 93, 783-88.

Fitch W.T., Giedd J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J Acoust Soc Am, 106, 1511-22.

Gelfer, Marylou P. & Mikos, Victoria A. (2005) The Relative Contributions of Speaking Fundamental Frequency and Formant Frequencies to Gender Identification Based on Isolated Vowels, Journal of Voice, 19, 544-554.

Gripper M., McBride M., Osafo-Yeboah B., Jiang X. (2007). Using the Callsign Acquisition Test (CAT) to compare the speech intelligibility of air versus bone conduction. International Journal of Industrial Ergonomics 37, 631-641.

Gripper, M.A., (2006). Evaluations of the callsign acquisition test (CAT) in acoustic environments. Unpublished Dissertation, North Carolina Agricultural and Technical State University.

Hain, T. (2003). Hearing Testing. [Online reference] Retrieved on November 04, 2008, from http://www.dizziness-and-balance.com/testing/audiogram.html

Handley, Z. (2008). Towards establishing a methodology for benchmarking speech synthesis for computer assisted language learning. [Online reference] Retrieved on November 07, 2008 from http://www.lsri.nottingham.ac.uk/zh/Papers/CLUK04.pdf

Henry, P & Mermagen, T. (2004). Use of bone conduction transmission for communication inside a military vehicle. Proceeding of the NATO Vehicle Habitability Conference, pp. 14-1 to 14-10. Prague (Czech Republic): October 4-6 NATO.

House, A., Williams, C., Hecker, M., & Kryter, K. (1965). Articulation testing methods: Consonantal differentiation with a closed-response set. Journal of Acoustical Society of America, 37,158-166.

Jenkins, J. S. (1998) The voice of the castrato,The Lancet 351, 1877 - 1880.

Kamm, C., Walker, M. & Rabiner, L. (1997). The role of speech processing in human computer intelligent communication. Paper presented at NSF Workshop on human-centered systems: Information, interactivity, and intelligence, February, 1997.

Kent, R. D. (1992). Intelligibility in speech disorders: Theory, measurement, and management. Philadelphia, PA: John Benjamins Publishing

Klatt D. H., (1983). Timing rules in Klattalk: Implications for models of speech production, 1, Acoust. SOC. Amer., Suppl. 1, vol. 73, p. 566.

Konkle, D., & Rintelmann, W. (1983). Principles of speech audiometry. Baltimore, MD: University Park Press.

Lee, E.-J., Nass, C., Brave, S., (2000). Can computer-generated speech have gender? An experimental test of gender stereotype. In: CHI Extended Abstract.

ACM Press, New York, pp. 289-290.

Letowski, T., Karsh, R., Vause, N., Shilling, R., Ballas, J., Brungart, D., & McKinley, R. (2001). Human factors military lexicon: Auditory displays [unpublished technical report]. U.S. Army Research Laboratory, Human Research and Engineering Directorate. Aberdeen Proving Grounds, MD.

Logan J., Greene B., Pisoni D. (1989). Segmental Intelligibility of Synthetic Speech Produced by Rule. Journal of the Acoustical Society of America, JASA vol. 86: 566-581.

MacFarlan, D. (1945). Speech hearing tests. Laryngscope, 55,71-115.

Markham, D. & Hazan, V. (2004) Acoustic-phonetic correlates of talker intelligibility in adults and children. Journal of the Acoustical Society of America, 116, 3108-3118.

Meyer Sound Labs (2008). Statistical measures of speech intelligibility. [Online reference] Retrieved May 21, 2008, from
http://www.meyersound.com/support/papers/speech/section3.htm

Nass C., Robles E., Bienenstock H., Treinen M., & Heenan C. (2003). Voice-based disclosure systems: Effects of modality, gender of prompt, and gender of user. International Journal of Speech Technology, 6(2):113-121, 2003.

Nass, C. & Lee, K.M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of Experimental Psychology: Applied, 7 (3), 171-181.

Nixon C. W., Morris L. J., McCavitt A. R., McKinley R. L., Anderson T. R. and McDaniel M. P. et al. (1998). Female communications in high levels of aircraft cockpit noises-part I: spectra, levels, and microphones, Aviation Space and Environmental Medicine 69 (1998), pp. 675-683.

Pisoni D.B., Nusbaum H.C. & Greene B.G. (1985). Perception of synthetic speech generated by rule. Proceedings of the 1EEE 73:1665-1676.

Rao, M. D. & Letowski, T. (2006). Callsign Acquisition Test (CAT): speech intelligibility in noise. Ear and Hearing 2006; 27:120-128.

Schroeter, J. (2001). The fundamentals of text-to-speech synthesis. [Online reference] Retrieved on November 02, 2008 from
http://www.voicexml.org/Review/Mar2001/features/tts.html

Silverstein B., Bilger R.C., Hanley T.D., & Steer M.D., (1953), "The relative intelligibility of male and female talkers", J. of Educational Psychology 44, 1953, 418-428.

Smither J.A. (1993). Short term memory demands in processing synthetic speech by young and old adults. Behavior and Information Technology 12, pp. 330-335.

Stevens C., Leesa N., Vonwillerb J., & Burnham D., (2005). On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference. Computer Speech and Language, 19, 129-146.

Studebaker, G.A. (1985). A "Rationalized" Arcsine Transform. Journal of Speech and Hearing Research, 28, 455-462.

Sydral, A., Bennett R., & Greenspan, S. (1994). Applied Speech Technology. Washington, DC: CRC Press, LLC

van Santen JPH, Pols LCW, Abe, M., Kahn, D., Keller, E., Vonwiller J., (1998). Report on the 3rd ESCA TTS workshop evaluation procedure. In: Proc. of the Third ESCA Workshop on Speech Synthesis, Jenolan Caves, Australia

Wertheim, E. (2008). The importance of effective communication. [Online reference] Retrieved on April 30, 2008, from

http://web.cba.neu.edu/∼ewertheim/interper/commun.htm